# KC-GenRe: A Knowledge-constrained Generative Re-ranking Method Based on Large Language Models for Knowledge Graph Completion

Yilin Wang

# CONTENTS

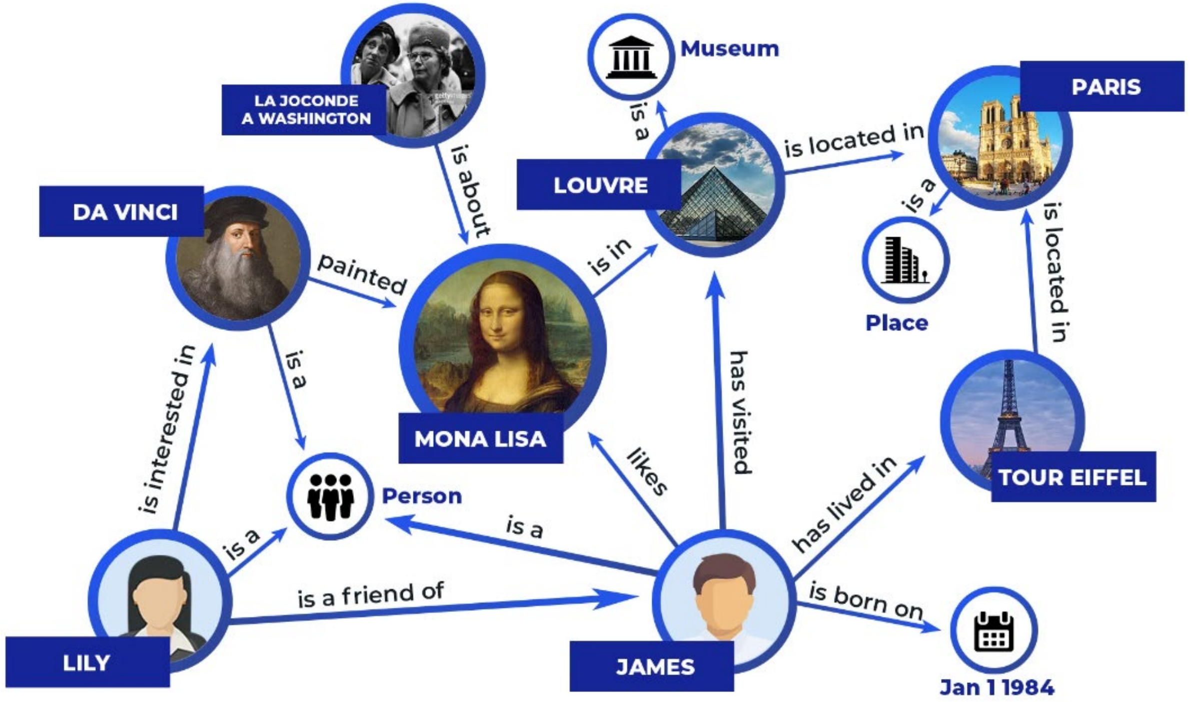
Part.01

# Background

# knowledge graph (KG)

- stores facts in the form of triples
  - Triple $(e_h, r, e_t)$: (head entity, relation, tail entity)

- Knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$
  - entity set $\mathcal{E}$
  - relation set $\mathcal{R}$
  - triple set $\mathcal{F}$



# knowledge graph completion (KGC)

- knowledge graphs are incomplete
  - Freebase: approximately 71% of the persons lack information on their birthplace; 75% of the persons lack information on their nationality.

- knowledge graph completion (KGC)
  - predict missing facts in the KG

Part.02

# Motivation

## motivation

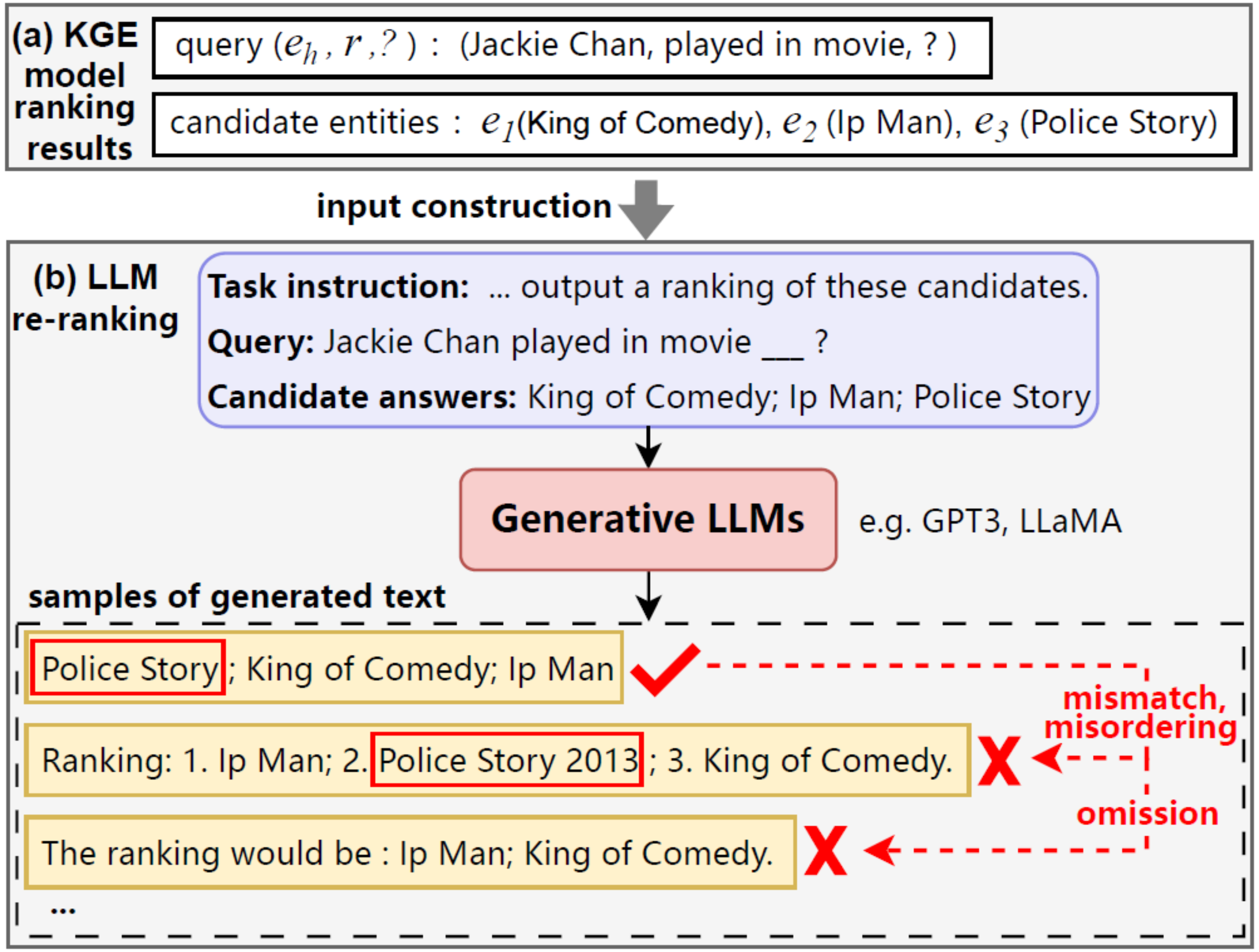
Generative PLM for KGC：

- **focus on ranking stage** to generate texts of the missing entity
- may require **manual assistance** to **match** the output texts with entities in KG

# motivation

Generative PLM for KGC **re-ranking**:

- **mismatch**: the generated text contains the candidate entity in KG, but in different textual.
- **misordering**: the correct answer is not predicted in the first position.
- **omission**: the output text fails to include all candidates, particularly the target one.



Part.03

# Method

# 1. The proposed method

- KC-GenRe: a knowledge-constrained generative reranking method

1 **mismatch** → outputting the order of option identifiers corresponding to these candidates

2 **misordering** → knowledge-guided interaction training method

3 **omission** → knowledge-augmented constrained inference method

## 1. The proposed method

**Contributions:**

- We propose KC-GenRe, a novel knowledge-constrained generative re-ranking model, which is **the first to utilize generative LLM for KGC re-ranking** as far as we know.
- We design **knowledge-guided interactive training** and **knowledge-augmented constrained inference** methods to stimulate the potential of generative LLMs and generate valid ranking of candidates for KGC.
- Experiments on four datasets show that **KG-GenRe outperforms start-of-arts results**, and extensive analysis demonstrates the effectiveness of the proposed components. Datasets and codes have been open sourced.

## 2. Overview of KC-GenRe

- **(1) Generative Re-ranking Formulation**
- **(2) Knowledge-guided Interactive Training**
- **(3) Knowledge-augmented Constrained Inference**

## 2. Overview of KC-GenRe

- **(1) Generative Re-ranking Formulation**

Query: $(e_h, r, ?)$

- First stage (ranking): KGE methods
  - top-K predicted candidate entities $E_c = \{e_{t_1}, e_{t_2}, \ldots, e_{t_K}\}$
  - corresponding scores $S_c = \{s_1, s_2, \ldots, s_K\}$
- Second stage (re-ranking): KC-GenRe
  - Inputs: $x_{in} = T_{in}(x_q, x_c)$

instruction template $T_{in}$, query sequence $x_q$, candidates sequence $x_c$

**Task instruction:** ... output a ranking of these candidates.
**Query:** Jackie Chan played in movie ___ ?
**Candidate answers:**
A. Ip Man ?
B. King of Comedy ?
C. Police Story ?

## 2. Overview of KC-GenRe

- **(1) Generative Re-ranking Formulation**

  Query: $(e_h, r, ?)$

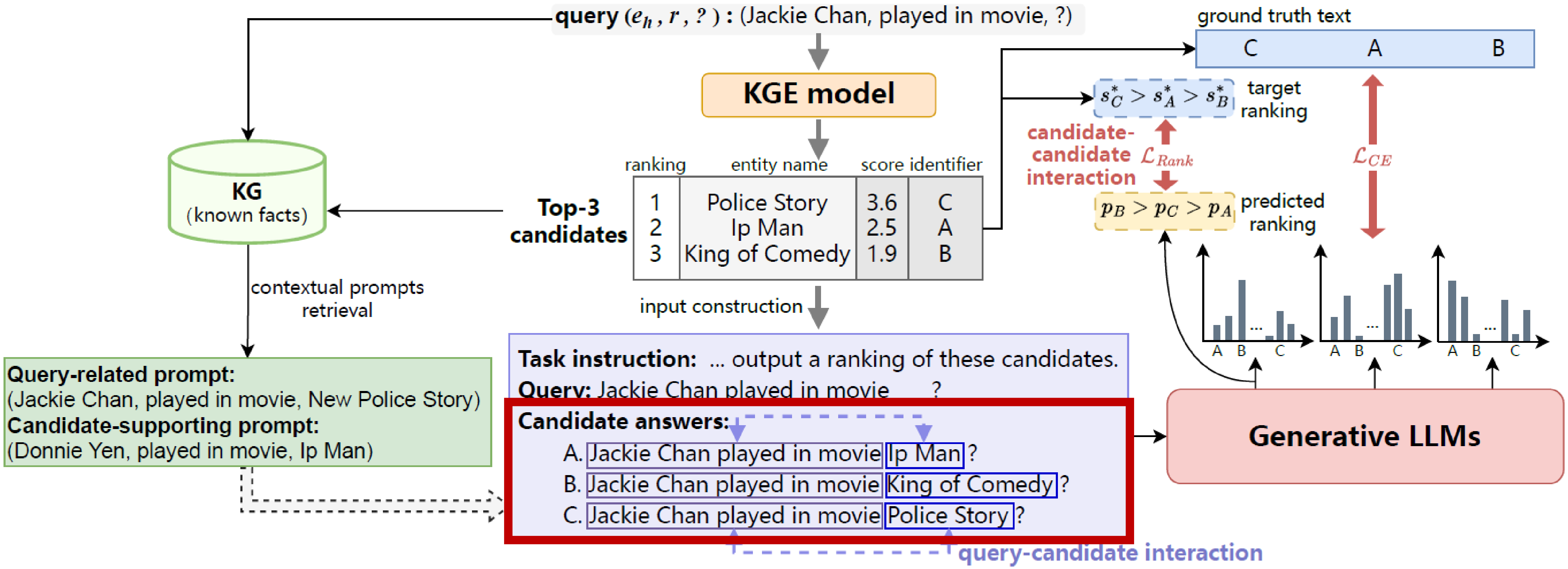
  - First stage (ranking): KGE methods
    - top-K predicted candidate entities $E_c = \{e_{t_1}, e_{t_2}, \ldots, e_{t_K}\}$
    - corresponding scores $S_c = \{s_1, s_2, \ldots, s_K\}$
  - Second stage (re-ranking): KC-GenRe
    - Outputs: ranking of option identifiers associated with candidates (instead of their labeled name texts)

**Task instruction:** ... output a ranking of these candidates.
**Query:** Jackie Chan played in movie ___ ?
**Candidate answers:**
A. Ip Man ?
B. King of Comedy ?
C. Police Story ?

→ **C A B**

# 2. Overview of KC-GenRe

- **(2) Knowledge-guided Interactive Training**
  - **query-candidate interaction**
  - candidate-candidate interaction

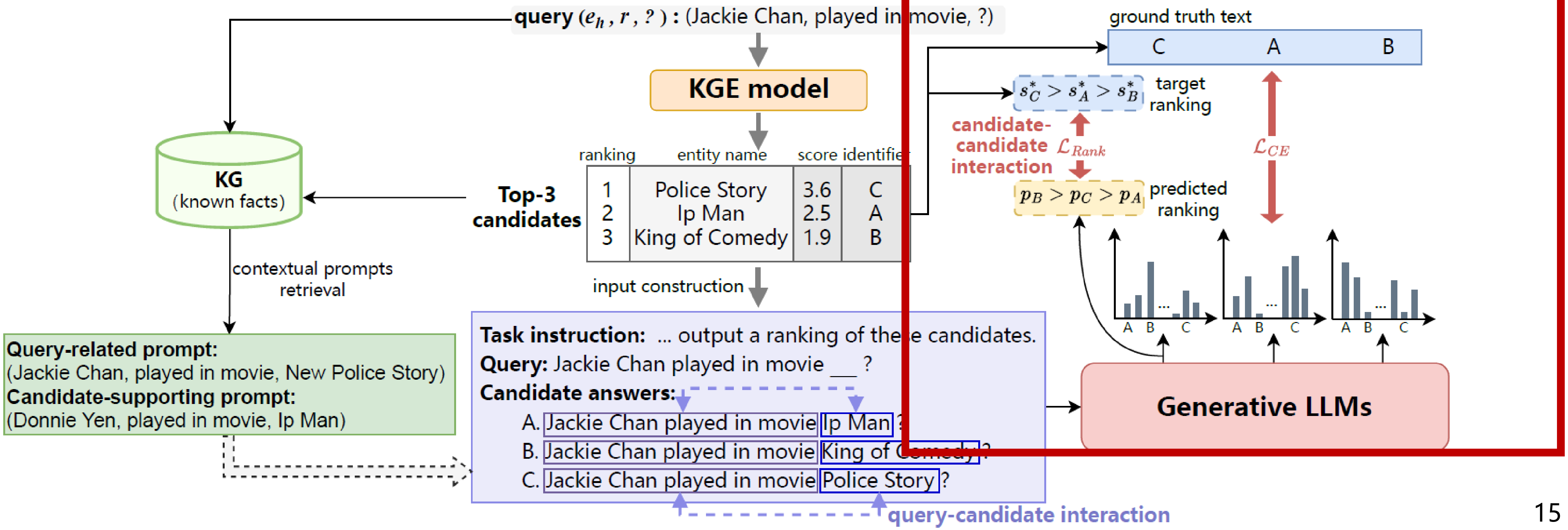


# 2. Overview of KC-GenRe

- **(2) Knowledge-guided Interactive Training**
  - query-candidate interaction
  - **candidate-candidate interaction**

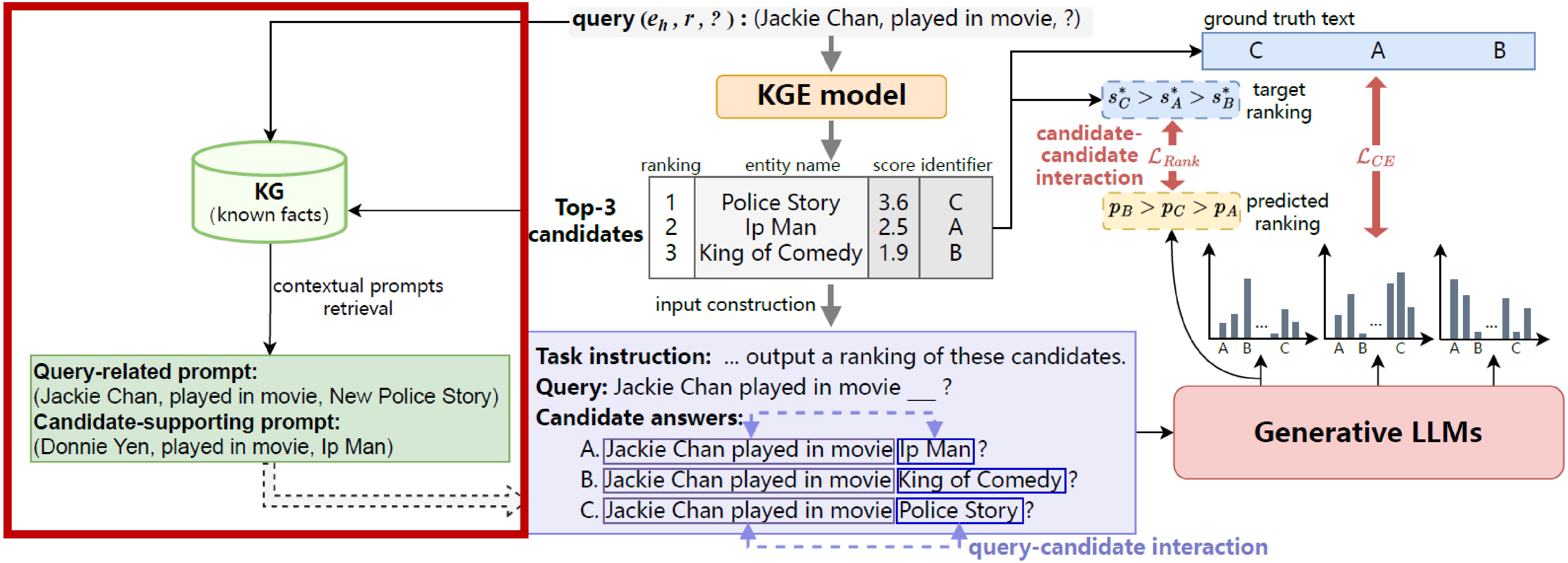
$$\mathcal{L}_{Rank} = \frac{C}{K^2} \sum_{s_i^* < s_j^*} \max(0, p_i - p_j), i, j \in \{1, \ldots, K\}$$

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{Rank}$$



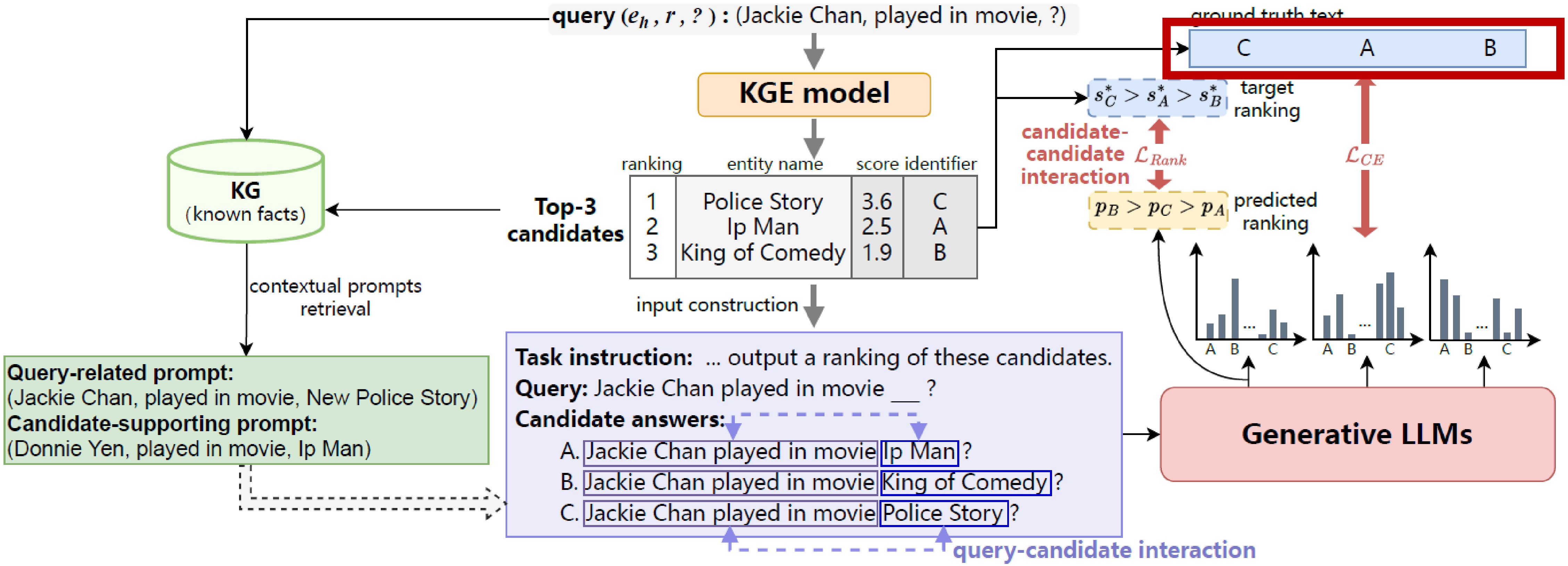
## 2. Overview of KC-GenRe

- **(3) Knowledge-augmented Constrained Inference**
  - **query-related prompt**
  - **candidate-supporting prompt**
  - constrained option generation



## 2. Overview of KC-GenRe

- **(3) Knowledge-augmented Constrained Inference**
  - query-related prompt
  - candidate-supporting prompt
  - **constrained option generation**



Part.04

# Experiments

## 1. Datasets

- Two curated KGs：Wiki27K，FB15K-237-N
- Two open KGs：ReVerb20K，ReVerb45K

| Dataset | $\lvert\mathcal{E}\rvert$ | $\lvert\mathcal{R}\rvert$ | # Train | # Valid | # Test |
|---|---|---|---|---|---|
| Wiki27K | 27,122 | 62 | 74,793 | 10,121 | 10,122 |
| FB15K-237-N | 13,104 | 93 | 87,282 | 7,041 | 8,226 |
| ReVerb20K | 11,065 | 11,058 | 15,499 | 1,550 | 2,325 |
| ReVerb45K | 27,008 | 21,623 | 35,970 | 3,598 | 5,395 |

Table 1: Dataset statistics.

## 2. Settings

- Hyperparameters
- Instruction templates
- Hardware: one A800 80G GPU

| Dataset | Wiki27K | FB15K-237-N | ReVerb20K | ReVerb45K |
|---|---|---|---|---|
| $K$ | 20 | 20 | 30 | 30 |
| $\lambda$ | 0.1 | 0.1 | 0.3 | 1.0 |

Table 2: Values of Hyperparameters.

| | |
|---|---|
| $T_{in}$ | Below is an instruction that describes a task, paired with a question and corresponding candidate answers. The questions and candidate answers have been combined into candidate corresponding statements. Combine what you know, output a ranking of these candidate answers.\n\n ### Question: $\{x_q\}$\n\n $\{x_c\}$ ### Response: |
| $T_{in}^k$ | Below is an instruction that describes a task, paired with a question and corresponding candidate answers. The questions and candidate answers have been combined into candidate corresponding statements. Knowledge related to some candidates will be provided that may be useful for ranking. Combine what you know and the following knowledge, output a ranking of these candidate answers.\n\n ### Supporting information: $\{x_q^k\}$\n\n ### Candidate supporting knowledge: $\{x_c^k\}$\n\n ### Question: $\{x_q\}$\n\n $\{x_c\}$ ### Response: |

Table 3: Instruction templates of KC-GenRe, where $x_q$, $x_c$, $x_q^k$ and $x_c^k$ represent query sequence, candidate sequence, query-related prompt, and candidate-supporting prompt respectively.

# 3. Main results

- KC-GenRe gains absolute improvements of 3.2% and **6.7%** for **MRR**, 4.4% and **7.7%** for **Hits@1** on Wiki27K and FB15K-237-N compared to PKGC.

| Model | Wiki27K | | | | FB15K-237-N | | | |
|---|---|---|---|---|---|---|---|---|
| | MRR | Hits@1 | Hits@3 | Hits@10 | MRR | Hits@1 | Hits@3 | Hits@10 |
| TransE† (Bordes et al., 2013) | 0.155 | 0.032 | 0.228 | 0.378 | 0.255 | 0.152 | 0.301 | 0.459 |
| TransC† (Lv et al., 2018) | 0.175 | 0.124 | 0.215 | 0.339 | 0.233 | 0.129 | 0.298 | 0.395 |
| ConvE† (Dettmers et al., 2018) | 0.226 | 0.164 | 0.244 | 0.354 | 0.273 | 0.192 | 0.305 | 0.429 |
| WWV† (Veira et al., 2019) | 0.198 | 0.157 | 0.237 | 0.365 | 0.269 | 0.137 | 0.287 | 0.443 |
| TuckER (Balazevic et al., 2019) | 0.249 | 0.185 | 0.269 | 0.385 | 0.309 | 0.227 | 0.340 | 0.474 |
| RotatE† (Sun et al., 2019) | 0.216 | 0.123 | 0.256 | 0.394 | 0.279 | 0.177 | 0.320 | 0.481 |
| KG-BERT† (Yao et al., 2019) | 0.192 | 0.119 | 0.219 | 0.352 | 0.203 | 0.139 | 0.201 | 0.403 |
| LP-RP-RR† (Kim et al., 2020) | 0.217 | 0.138 | 0.235 | 0.379 | 0.248 | 0.155 | 0.256 | 0.436 |
| PKGC† (Lv et al., 2022) | <u>0.285</u> | <u>0.230</u> | <u>0.305</u> | **0.409** | <u>0.332</u> | <u>0.261</u> | <u>0.346</u> | <u>0.487</u> |
| KC-GenRe | **0.317** | **0.274** | **0.330** | <u>0.408</u> | **0.399** | **0.338** | **0.427** | **0.505** |

Table 4: Link prediction results on two curated KGs. Best results are in bold and second best are underlined. [†]: results are taken from PKGC (Lv et al., 2022).

# 3. Main results

- KC-GenRe achieves higher performance than previous methods with improvements of 2.1% and **3.5%** for **MRR**, 2.1% and **3.8%** for **Hits@1** on ReVerb20K and ReVerb45K, respectively.

| Model | ReVerb20K | | | | | ReVerb45K | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MRR | MR | Hits@1 | Hits@3 | Hits@10 | MRR | MR | Hits@1 | Hits@3 | Hits@10 |
| TransE (Bordes et al., 2013) | 0.138 | 1150.5 | 0.034 | 0.201 | 0.316 | 0.202 | 1889.5 | 0.122 | 0.243 | 0.346 |
| ComplEx (Trouillon et al., 2016) | 0.038 | 4486.5 | 0.017 | 0.043 | 0.071 | 0.068 | 5659.8 | 0.054 | 0.071 | 0.093 |
| R-GCN (Schlichtkrull et al., 2018) | 0.122 | 1204.3 | - | - | 0.187 | 0.042 | 2866.8 | - | - | 0.046 |
| ConvE (Dettmers et al., 2018) | 0.262 | 1483.7 | 0.203 | 0.287 | 0.371 | 0.218 | 3306.8 | 0.166 | 0.243 | 0.314 |
| KG-BERT (Yao et al., 2019) | 0.047 | 420.4 | 0.014 | 0.039 | 0.105 | 0.123 | 1325.8 | 0.070 | 0.131 | 0.223 |
| RotatE (Sun et al., 2019) | 0.065 | 2861.5 | 0.043 | 0.069 | 0.108 | 0.141 | 3033.4 | 0.110 | 0.147 | 0.196 |
| PairRE (Chao et al., 2021) | 0.213 | 1366.2 | 0.166 | 0.229 | 0.296 | 0.205 | 2608.4 | 0.153 | 0.228 | 0.302 |
| ResNet (Lovelace et al., 2021) | 0.224 | 2258.4 | 0.188 | 0.240 | 0.292 | 0.181 | 3928.9 | 0.150 | 0.196 | 0.242 |
| BertResNet-ReRank (Lovelace et al., 2021) | 0.272 | 1245.6 | 0.225 | 0.294 | 0.347 | 0.208 | 2773.4 | 0.166 | 0.227 | 0.281 |
| CaRe (Gupta et al., 2019) | 0.318 | 973.2 | - | - | 0.439 | 0.324 | 1308.0 | - | - | 0.456 |
| OKGIT (Chandrahas and Talukdar, 2021) | 0.359 | 527.1 | 0.282 | 0.394 | 0.499 | 0.332 | <u>773.9</u> | 0.261 | 0.363 | 0.464 |
| OKGSE (Xie et al., 2022a) | 0.372 | 487.3 | 0.291 | 0.408 | <u>0.524</u> | 0.342 | **771.1** | 0.274 | 0.371 | 0.473 |
| CEKFA (Wang et al., 2023b) | <u>0.387</u> | <u>416.7</u> | <u>0.310</u> | <u>0.427</u> | 0.515 | <u>0.369</u> | 884.5 | <u>0.294</u> | <u>0.409</u> | <u>0.502</u> |
| KC-GenRe | **0.408** | **410.8** | **0.331** | **0.450** | **0.547** | **0.404** | 874.1 | **0.332** | **0.444** | **0.534** |

Table 5: Link prediction results on two open KGs. Best results are in bold and second best are underlined.

# 4. Ablation study

- Compared to **Base (KGE model without re-ranking)** , KC-GenRe gains of 4.1%, 5.2%, 6.8% and **9.0%** in MRR on ReVerb20K, ReVerb45K, Wiki27K and **FB15K-237-N** respectively

Reranking with LLM

| | QCI | CCI | QP | CP | CG | ReVerb20K MRR | ReVerb20K Hits@1 | ReVerb45K MRR | ReVerb45K Hits@1 |
|---|---|---|---|---|---|---|---|---|---|
| Base | | | | | | 0.367 | 0.288 | 0.352 | 0.274 |
| 1 | | | | | | 0.350 | 0.263 | 0.334 | 0.246 |
| 2 | | ✓ | ✓ | ✓ | ✓ | 0.383 | 0.304 | 0.365 | 0.282 |
| 3 | ✓ | | ✓ | ✓ | ✓ | 0.405 | 0.326 | 0.381 | 0.298 |
| 4 | ✓ | ✓ | | ✓ | ✓ | 0.399 | 0.320 | 0.397 | 0.323 |
| 5 | ✓ | ✓ | ✓ | | ✓ | 0.403 | 0.328 | 0.400 | 0.327 |
| 6 | ✓ | ✓ | | | ✓ | 0.370 | 0.284 | 0.360 | 0.271 |
| 7 | ✓ | ✓ | ✓ | ✓ | | 0.367 | 0.289 | 0.352 | 0.273 |
| KC-GenRe | ✓ | ✓ | ✓ | ✓ | ✓ | 0.408 | 0.331 | 0.404 | 0.332 |

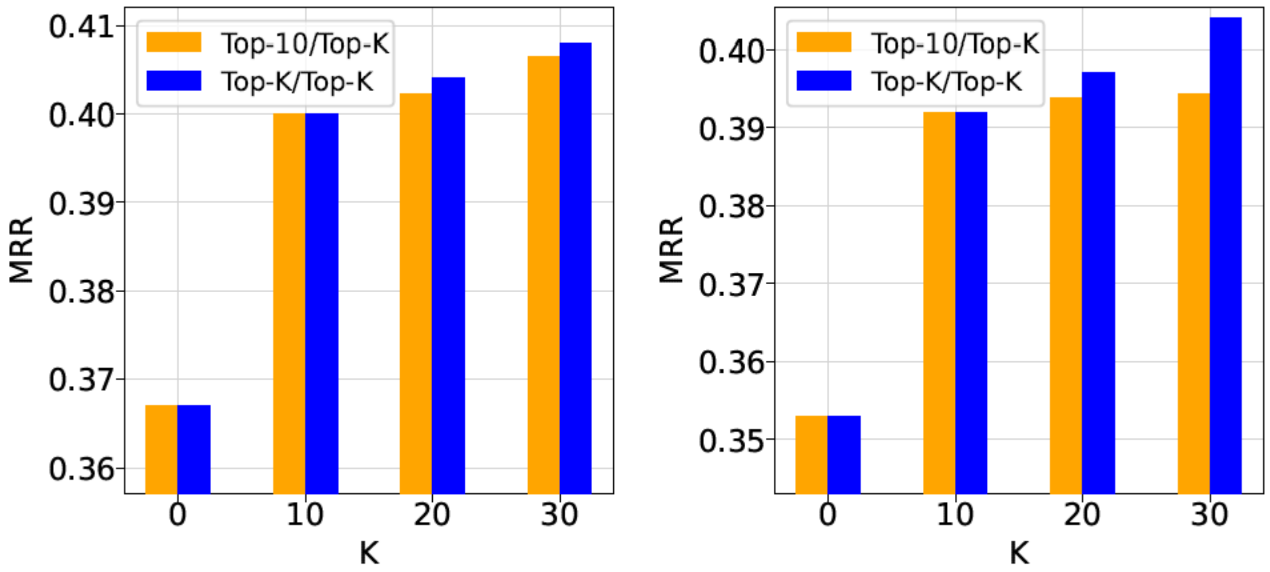
Table 6: Ablation results on open KGs, where "Base" is CEKFA-KFARe (Wang et al., 2023b).

| | QCI | CCI | DP | CG | Wiki27K MRR | Wiki27K Hits@1 | FB15K-237-N MRR | FB15K-237-N Hits@1 |
|---|---|---|---|---|---|---|---|---|
| Base | | | | | 0.249 | 0.185 | 0.309 | 0.227 |
| 1 | | | | | 0.283 | 0.227 | 0.329 | 0.248 |
| 2 | | ✓ | ✓ | ✓ | 0.314 | 0.268 | 0.340 | 0.257 |
| 3 | ✓ | | ✓ | ✓ | 0.311 | 0.266 | 0.361 | 0.284 |
| 4 | ✓ | ✓ | | ✓ | 0.298 | 0.249 | 0.353 | 0.277 |
| 5 | ✓ | ✓ | ✓ | | 0.245 | 0.178 | 0.303 | 0.218 |
| KC-GenRe | ✓ | ✓ | ✓ | ✓ | 0.317 | 0.274 | 0.399 | 0.338 |

Table 7: Ablation results on curated KGs, where "Base" denotes TuckER (Balazevic et al., 2019) and "DP" represents entity definition prompt.

# 5. Effects of reranking number K

- The model trained to rank the top-K candidates has the ability to rank beyond the top-K.




(a) ReVerb20K (b) ReVerb45K

Figure 3: Effects of re-ranking number $K$. The front and back of the slash in the legend represent the values of $K$ during training and testing, respectively.

# 6. Effects of candidate-candidate interaction

- The prediction performance is able to be improved after applying candidate-candidate interaction.

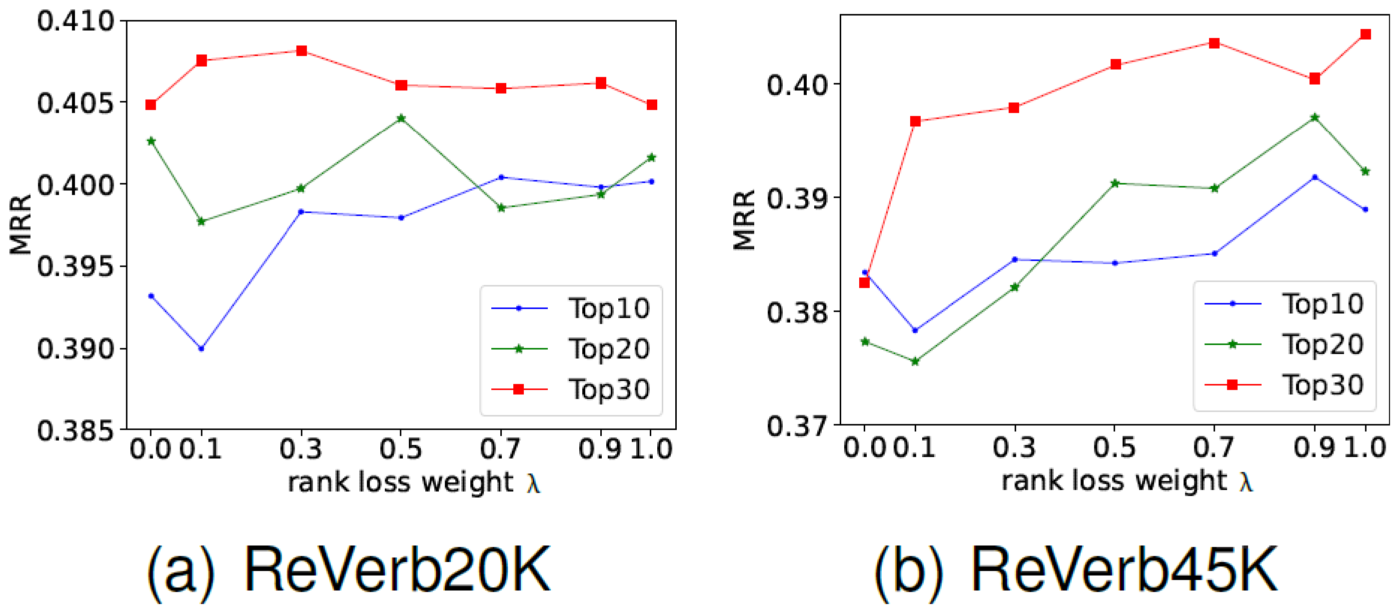



Figure 4: Influences of weight $\lambda$ in Eq.(2) with different re-ranking number $K$ (Top-$K$).

# 7. Effects of Different LLMs

- Various LLMs exhibit varying levels of performance (the lift may be slight).

| LLM | ReVerb20K | | ReVerb45K | |
|---|---|---|---|---|
| | MRR | Hits@1 | MRR | Hits@1 |
| LLaMA-7b | 0.400 | 0.324 | 0.392 | 0.325 |
| LLaMA2-7b | 0.406 | 0.334 | 0.397 | 0.329 |
| LLaMA-13b | 0.403 | 0.331 | 0.392 | 0.322 |
| LLaMA-65b | 0.400 | 0.326 | 0.404 | 0.342 |

Table 8: Link prediction results of KC-GenRe implemented with different LLMs when $K = 10$.

# 8. Case study

| | | |
|---|---|---|
| Example 2 | Query | Yugioh started in ___ ? |
| | Candidates | A. Kazuki Takahashi B. China C. Israel D. Jerusalem E. Lebanon F. Japan G. India H. Galilee<br>I. Saudi Arabia J. Asian Russia |
| | QP | Yugioh is property of Kazuki Takahashi;<br>Yu Gi is the property of Konami;<br>Judaism officially began in Yavneh. |
| | CP | Yugioh is property of Kazuki Takahashi. |
| | Output Ranking | F. A. G. C. E. B. I. D. H. J. |
| Example 3 | Query | Lyme disease is caused by ___ ? |
| | Candidates | A. lyme B. lyme disease C. inflammatory bowel disease D. lyme borreliosis E. diabetes<br>F. zithromax G. burgdorferi H. borrelia burgdorferi I. biaxin J. diabetes mellitus |
| | QP | borrelia burgdorferi is the causative agent of lyme disease;<br>borrelia burgdorferi is the etiological agent of lyme disease;<br>lyme disease does occur in California. |
| | CP | lyme opened the door for inflammatory bowel disease;<br>lyme disease is also called lyme borreliosis;<br>zithromax had a positive test for lyme;<br>zithromax is an excellent choice for treating lyme;<br>burgdorferi is responsible for lyme disease;<br>borrelia burgdorferi is the causative agent of lyme disease;<br>borrelia burgdorferi is the etiological agent of lyme disease;<br>biaxin is a natural remedy for lyme disease. |
| | Output Ranking | H. G. A. B. D. C. J. E. F. I. |

Table 10: Examples from datasets ReVerb45K, where QP and CP represent query-related prompt and candidate-supporting prompt respectively. The ground truth tail NPs, helpful information in QP and CP are highlighted. Note that CP may be empty, and different NPs may point to the same target entity, e.g. burgdorferi and borrelia burgdorferi in the third example.

# Thanks!